Stochastic Gradient Descent Learning in Text Classification

Chiemeziem Oguayo

Table of Contents

Problem Statement

The objective of this project is to optimize the accuracy of selected classifiers using SGD learning and to fine tune hyperparameters in order to enhance its performance

Description of Dataset

The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 180,000 attacks.

Geography: Worldwide

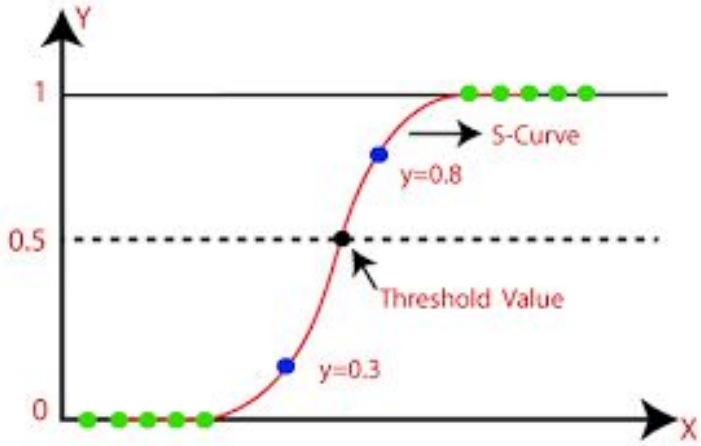
Time period: 1970-2017, *except 1993*

Unit of analysis: Attack

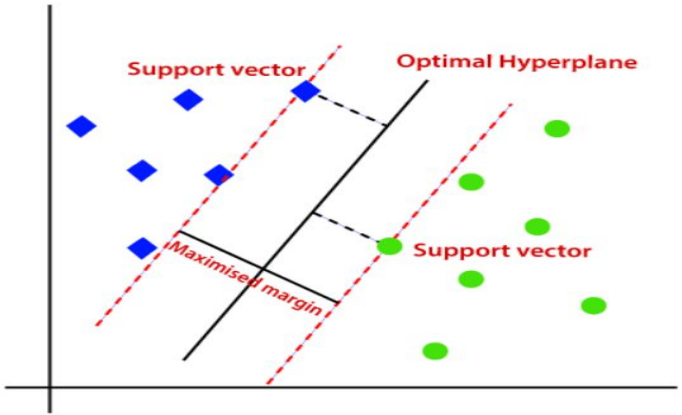
Variables: >100 variables on location, tactics, perpetrators, targets, and outcomes

Models

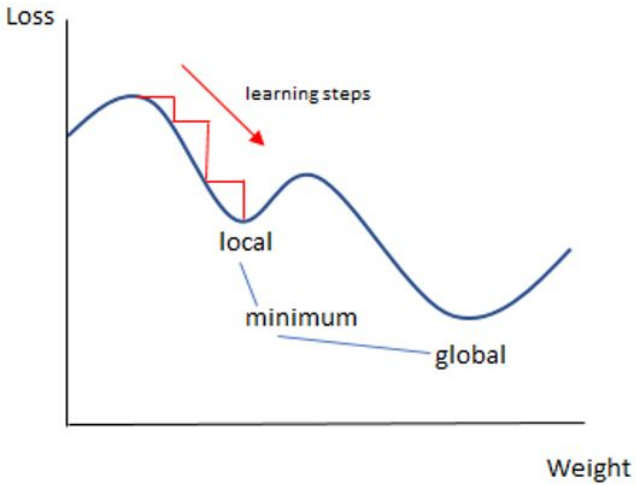
Logistic Regression



Support Vector Machine



SGD Learning/Classifier



- A linear classifier(SVM, Logistic regression, etc.) optimized by SGD
- Calculates the gradient for one observation picked at random

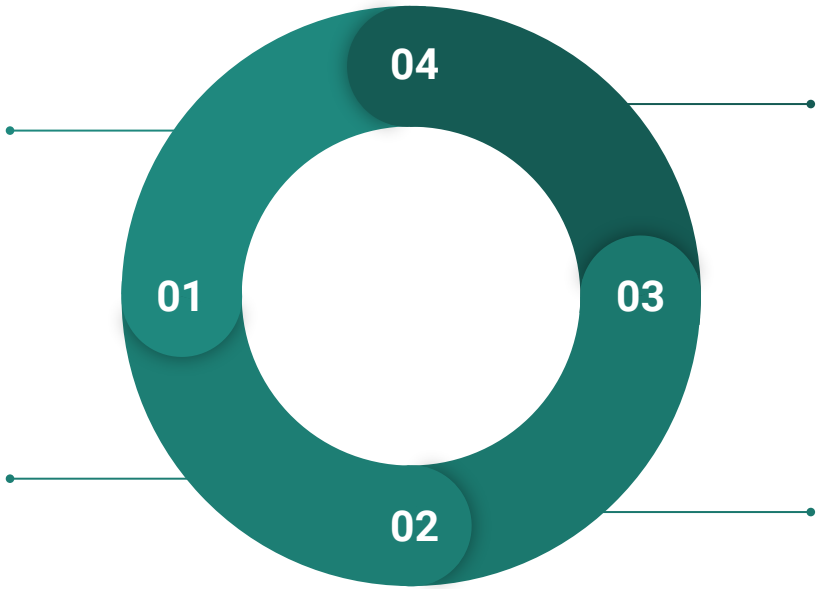
Experimental Setup

Data Preprocessing

Removing NA values, special characters, removing stopwords

Feature Extraction

TF-IDF, normalization, SMOTE



Validation & Evaluation

Accuracy, Precision, Recall, F1

Classification Design

- Applying default parameters
- Applying SGD Learning
- Applying SGD Learning with optimization /fine tuning

Classifier	Performance Metrics			
	Accuracy	Precision	Recall	F1
Logistic	0.876	0.809	0.629	0.685
SVM	0.859	0.817	0.525	0.573
Logistic + SGD	0.842	0.527	0.429	0.451
SVM + SGD	0.858	0.813	0.521	0.568

Gradient Descent

PENALTY: the aka regularization, two options investigated L1(abs value of weights) and L2(penalizes sum of squared weights).

ALPHA: a constant used to compute the step size and the learning rate.

Scaling the train set will cause data leakage if done before cross validation, because cross validation further divides a train set into additional train and test sets. It is recommended to use Pipeline with GridSearchCV

Conclusion

Explored different experiments to compare the performance of the classifiers without SGD learning, with SGD learning and with SGD learning including hyper-parameter discovery.

Can’t determine that SGD learning optimizes the accuracy of the selected classifiers

Limitations & Future Direction

- Gradient Descent(correct scaling)
- n-grams representation

QUESTIONS?